

A Probabilistic Logic Programming Approach to Automatic Video Montage[†]

B. Aerts¹ and T. Goedemé¹ and J. Vennekens¹

¹KU Leuven, Technology Campus De Nayer, Research Group EAVISE, Belgium, email: b.aerts, toon.goedeme, joost.vennekens@kuleuven.be

Abstract

Hiring a professional camera crew to cover an event such as a lecture, sports game or musical performance may be prohibitively expensive. The CAMETRON project aims at drastically reducing this cost by developing an (almost) fully automated system that can produce video recordings of such events with a quality similar to that of a professional crew. This system consists of different components, including intelligent Pan-Tilt-Zoom cameras and UAVs that act as “virtual camera men”. To combine the footage of these different cameras into a single coherent and pleasant-to-watch video, a “virtual editor” is needed.

Human editors typically follow a number of different—and sometimes contradictory—cinematographic “rules” to accomplish this task. To develop our virtual editor, we will follow a declarative approach, in which we explicitly represent these rules. This approach has the benefit that it offers a great deal of flexibility in deciding which rules should be taken into account and how they should take priority over each other. It also allows us to reuse the same knowledge to perform different tasks: we cannot only use the rules to generate a montage, but also to evaluate the quality of a given montage or to learn certain properties of good montages from given examples.

To represent the rules, we need a suitable knowledge representation language. A particular challenge is that cinematographic rules are not strict: they are guidelines that are typically followed, but not always. Indeed, the rules may sometimes contradict each other, and even if they do not, a human editor may still choose to ignore a rule, simply because the result “feels” better. A virtual editor should therefore not rigidly follow the rules, but it should sometimes deviate from them in order to give the montage a more interesting and natural flavour, thereby mimicking the creativity of a human editor. For this reason, we have chosen to make use of the Probabilistic Logic Programming language CP-logic and its implementation in the Problog system, which allows us to represent these rules in a non-deterministic way. This has the additional benefit that—just like a human editor—the system is able to produce different montages from the same input streams.

The proposed editing system takes as input a number of different video streams, together with a computer vision analysis of each of these streams. For each frame in each stream, we expect this analysis to provide information, such as the presence of people, the type of shot and the action the main subject performs.

The goal of our editing system is to decide for each point in time which of the available camera feeds will be used. Those decisions are made based on the cinematographic model described in the paper. The output of the system is the single video stream that is thus constructed.

We demonstrate that the resulting system is able to produce real-time edits of different video streams. In order to verify the quality of the resulting montage, we subjected the virtual editor to a “Turing test”: we asked 58 test subjects to distinguish between the output of our system and a professionally made montage of the same video streams. 31 subjects correctly identified the professionally edited clip. The difference between this outcome and one that could be produced by random guessing is not statistically significant.

We conclude that our editing system indeed provides a good approximation of the quality delivered by a professional editor for this particular case study of lecture recording.

[†] This paper originally appeared at the 22nd european conference on artificial intelligence (ECAI 2016)